# 確率・統計基礎: 確率論の初歩

森 立平

# 確率とは

- 世の中のランダムな事象を数学的に記述したもの
- 統計・機械学習の基礎となっている
- 極限定理: 確率 $p$ で表が出るコインを $N$ 回投げた時に、表が出る回数 $F$ は
  - 大数の法則: $\frac{F}{N} \to p$
  - 大偏差原理: $\Pr\left(\left|\frac{F}{N} - p\right| > \epsilon\right) = \mathrm{e}^{-\beta N}$
  - 中心極限定理: $\frac{F-pN}{\sqrt{N}} \to$ 正規分布

# 統計とは

確率論は統計の基礎となる

- この患者は病気ですか？
- 明日の降水量はいくつですか？

# 確率論の数学モデルとは？

測度論！

- 確率は面積 (= 測度) のようなもの。
- 数学的に厳密に様々な結果 (大数の法則、大偏差原理、中心極限定理) が証明できる。
- とても重要だが勉強するのは 3 年生以降 (解析学要論 II 、確率論)。
- この授業では測度論には深入りせず、基礎的な部分はごまかしながら確率・統計の重要な理論について学ぶ。

# 標本空間と確率

- $\Omega$： 集合。これの部分集合に確率が与えられる
- $P\colon 2^{\Omega} \to [0,1]$. 確率を与える関数

Example

- $\Omega = \{表, 裏\}$.
- $P(\emptyset) = 0$, $P(\{表\}) = P(\{裏\}) = \frac{1}{2}$, $P(\{表, 裏\}) = 1$.

- $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- $P(A) = \frac{|A|}{6} \quad \forall A \subseteq \Omega$.

- $\Omega = \{晴, 雨, 雪\}$.
- $P(\emptyset) = 0$, $P(\{晴\}) = 0.7$, $P(\{雨\}) = 0.2$, $P(\{雪\}) = 0.1$,
  $P(\{晴, 雨\}) = 0.9$, $P(\{雨, 雪\}) = 0.3$, $P(\{雪, 晴\}) = 0.8$
  $P(\{晴, 雨, 雪\}) = 1$.

# 確率論の公理有限版

- $\Omega$ : 有限集合 (標本空間)
- $P\colon 2^{\Omega} \to [0,1]$ (確率測度)

確率の公理

1. $P(\Omega) = 1$.
2. $\forall A, B \subseteq \Omega,\ A \cap B = \varnothing \implies P(A \cup B) = P(A) + P(B)$ (有限加法性).

$\Omega$ を無限集合にしたい場合はどうする？

# 確率論の公理可算版

- $\Omega$: 高々可算集合 (標本空間)
- $P: 2^\Omega \to [0,1]$ (確率測度)

確率の公理

1. $P(\Omega) = 1$.
2. $\forall A, B \subseteq \Omega,\ A \cap B = \varnothing \implies P(A \cup B) = P(A) + P(B)$ (有限加法性).

2'. $\forall (A_n \subseteq \Omega)_{n=1,\ldots}$,
$\forall i \neq j,\ A_i \cap A_j = \varnothing \implies P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$
(完全加法性, $\sigma$-加法性).

# なぜ有限加法性じゃ駄目なの？ ←駄目じゃないけど...

## Theorem (確率測度の連続性)

$P$ を完全加法確率測度とする。事象 $A_1 \subseteq A_2 \subseteq \cdots$ について

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n\to\infty} P(A_n).$$

## Proof.

$B_1 := A_1$, $i \geq 2$ について $B_i := A_i \setminus A_{i-1}$ とおく。このとき、$i \neq j$ について $B_i \cap B_j = \emptyset$。また、$\bigcup_{i=1}^{n} B_i = \bigcup_{i=1}^{n} A_i = A_n$.

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} A_i\right) &= P\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} P(B_i) \\ &= \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i) = \lim_{n\to\infty} P\left(\bigcup_{i=1}^{n} B_i\right) = \lim_{n\to\infty} P(A_n). \quad \square \end{aligned}$$

# 実数上の測度

- $\Omega = [0, 1)$.
- $P\colon 2^{\Omega} \to [0, 1]$

ここで

$$P([a, b]) = b - a$$

となるようにしたい。すると $P([a, a + \epsilon]) = \epsilon$ なので $P(\{a\}) \leq \epsilon$ である。任意の $\epsilon > 0$ について成り立つので $P(\{a\}) = 0$ である。

任意の 1 点の確率は 0 だけど全体の確率は 1 になるようにする。なので加法性は非可算には一般化しない。

# 確率論の公理最終版?

これでよいのか？

- $\Omega$ : (非可算でもよい) 集合 (標本空間)
- $P : 2^{\Omega} \to [0,1]$ (確率測度)

確率の公理

1. $P(\Omega) = 1$.
2. $\forall A, B \subseteq \Omega,\ A \cap B = \emptyset \implies P(A \cup B) = P(A) + P(B)$ (有限加法性).

2'. $\forall (A_n \subseteq \Omega)_{n=1,\dots}$,
$\forall i \neq j,\ A_i \cap A_j = \emptyset \implies P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$
(完全加法性, $\sigma$-加法性).

これで大丈夫？

# ダメー (完全加法性+選択公理)

- $\Omega = [0, 1)$.
- $\forall c \in \Omega,\ A \subseteq \Omega,\ P(A + c) = P(A)$ where
  $A + c := \{a + c - \lfloor a + c \rfloor \mid a \in A\}$.

選択公理を仮定するとそのような確率空間は存在しない。

$\Omega$ の上の同値関係を $x \sim y \overset{\text{def}}{\iff} x - y \in \mathbb{Q}$ と定義する。

$\Omega$ の上の同値類から一つずつ要素を選んで集合 $V$ を作る (選択公理)。

$$
\begin{aligned}
1 = P([0,1)) &= P\left(\bigcup_{x \in \mathbb{Q}} (V + x)\right) \\
&= \sum_{x \in \mathbb{Q}} P(V + x) \qquad \text{(完全加法性)} \\
&= \sum_{x \in \mathbb{Q}} P(V)
\end{aligned}
$$

同じ値を無限回足して 1 にはできない。

# ちゃんとした確率論の公理

確率空間 $(\Omega, \mathcal{F}, P)$

- $\Omega$： 集合 (標本空間)
- $\mathcal{F} \subseteq 2^{\Omega}$ (事象の集合、可測集合)
- $P : \mathcal{F} \to [0,1]$ (確率測度、確率は $\mathcal{F}$ の要素にのみ定義される！)

可測集合の公理 ($\sigma$-加法族、完全加法族)

1. $\emptyset \in \mathcal{F}$.
2. $A \in \mathcal{F} \implies A^c \in \mathcal{F}$.

3'. $(A_n \in \mathcal{F})_{n=1,\dots} \implies \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ (完全加法性, $\sigma$-加法性).

確率の公理

1. $P(\Omega) = 1$.

2'. $\forall (A_n \in \mathcal{F})_{n=1,\dots}$,
$\forall i \neq j, A_i \cap A_j = \emptyset \implies P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$
(完全加法性, $\sigma$-加法性).

# この授業の方針

ここまでのまとめ

- $\Omega$ が高々可算の場合は $\mathcal{F} = 2^{\Omega}$ とできる。
- $\Omega$ が非可算の場合は (平行移動で不変などの条件を課すと) すべての部分集合を可測とできない場合がある (要選択公理)。

この授業では $\mathcal{F} = 2^{\Omega}$ のつもりですすめる。
$\Omega$ の部分集合を事象と呼ぶ。

実際に考えるような確率空間では非可測な集合 $A \notin \mathcal{F}$ は選択公理を使わないと構成できない。

普通に考える集合は全部可測。

この授業は測度論の授業じゃないので、非可測な集合については考えないことにする。

今後、確率空間は $(\Omega, \mathcal{F}, P)$ じゃなくて $(\Omega, P)$ とする。

# 無限個の集合の操作

集合族 $(A_i \subseteq \Omega)_{i=1,2,\dots}$ について、

$$\bigcup_{i=1}^{\infty} A_i = \{\omega \in \Omega \mid \exists i \in \{1, 2, \dots\} \quad \omega \in A_i\}$$

$$\bigcap_{i=1}^{\infty} A_i = \{\omega \in \Omega \mid \forall i \in \{1, 2, \dots\} \quad \omega \in A_i\}$$

**Lemma (ド・モルガンの法則)**

$$\left(\bigcap_{i=1}^{\infty} A_i\right)^c = \bigcup_{i=1}^{\infty} A_i^c, \qquad \left(\bigcup_{i=1}^{\infty} A_i\right)^c = \bigcap_{i=1}^{\infty} A_i^c$$

*Proof.*

$$\left(\bigcup_{i=1}^{\infty} A_i\right)^c = \{\omega \in \Omega \mid \forall i \in \{1, 2, \dots\} \quad \omega \in \Omega \setminus A_i\} = \bigcap_{i=1}^{\infty} A_i^c.$$

□

# 確率の性質

任意の $A, B \in 2^{\Omega}$ について，

- $P(\Omega) = 1$, $P(\emptyset) = 0$.
- $P(A^c) = 1 - P(A) \quad \because 1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$.
- $B \subseteq A \implies P(B) \leq P(A)$
  $\because P(A) = P(B \cup (A \setminus B)) = P(B) + P(A \setminus B) \geq P(B)$.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  $\because P(A \cup B) = P(B) + P(A \setminus B)$, $P(A) = P(A \setminus B) + P(A \cap B)$
- $P(A \cup B) \leq P(A) + P(B)$ (ブールの不等式, union bound).

# 条件付き確率

事象 $B \in 2^\Omega$ が起きる条件下で事象 $A \in 2^\Omega$ の条件付き確率を

$$P(A \mid B) := \frac{P(A \cap B)}{P(B)}$$

と定義する。ただし、$P(B) = 0$ となる $B \in 2^\Omega$ については定義されない。

$(\Omega, P(\cdot \mid B))$ は確率空間となる。

1. $P(\Omega \mid B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$

2'. $\forall (A_n \in 2^\Omega)_{n=1,\dots},\ \forall i \neq j,\ A_i \cap A_j = \varnothing \implies$

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} A_n \mid B\right) &= \frac{P\left(\left(\bigcup_{n=1}^{\infty} A_n\right) \cap B\right)}{P(B)} \\ &= \frac{P\left(\bigcup_{n=1}^{\infty} (A_n \cap B)\right)}{P(B)} \\ &= \frac{\sum_{n=1}^{\infty} P(A_n \cap B)}{P(B)} = \sum_{n=1}^{\infty} P(A_n \mid B) \end{aligned}$$

また、$P(A \cap B) = P(A)P(B)$ もしくは等価である $P(A \mid B) = P(A)$ のとき、事象 $A$ と $B$ は独立であるという。

# 確率変数

確率空間 $(\Omega, P)$ について、標本空間 $\Omega$ から $\mathbb{R}$ への写像 $X\colon \Omega \to \mathbb{R}$ を確率変数という。

任意の $S \subseteq \mathbb{R}$ について

$$\Pr(X \in S) := P(X^{-1}(S)) = P(\{\omega \in \Omega \mid X(\omega) \in S\})$$

と定義する。

その他によく使う記法として、任意の $a \in \mathbb{R}$ について

$$\Pr(X \geq a) := \Pr(X \in \{x \in \mathbb{R} \mid x \geq a\}) = \Pr(\{\omega \in \Omega \mid X(\omega) \geq a)$$

同様に $\Pr(X > a), \Pr(X \leq a), \Pr(X < a), \Pr(X = a)$ なども定義される。

# 確率質量関数

- 特に $X$ の像が高々可算のとき、$X$ を離散型確率変数という。
- 離散型確率変数でない確率変数を連続型確率変数という。
- 離散型確率変数 $X$ について $f_X(a) := \Pr(X = a)$ を確率質量関数という。
- 確率変数 $X$ について $F_X(a) := \Pr(X \leq a)$ を累積分布関数という。

# 離散型確率変数の例 (有限の場合)

Example

- $\Omega = \{\mathrm{H}, \mathrm{T}\}$.
- $P(\{\mathrm{H}\}) = P(\{\mathrm{T}\}) = 1/2$.

確率変数 $X\colon \Omega \to \mathbb{R}$ を $X(\mathrm{H}) = 1$, $X(\mathrm{T}) = 0$ と定義する。
このとき、$\Pr(X = 1) = P(X^{-1}(1)) = P(\{\mathrm{H}\}) = 1/2$ である。

Example

- $\Omega = \{\mathrm{H}, \mathrm{T}\}^2 = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$.
- $P(\{\mathrm{HH}\}) = P(\{\mathrm{HT}\}) = P(\{\mathrm{TH}\}) = P(\{\mathrm{TT}\}) = 1/4$.

確率変数 $X\colon \Omega \to \mathbb{R}$ を $X(\mathrm{HH}) = 2$, $X(\mathrm{HT}) = X(\mathrm{TH}) = 1$, $X(\mathrm{TT}) = 0$ と定義する。
このとき、
$\Pr(X \geq 1) = P(\{\omega \in \Omega \mid X(\omega) \geq 1) = P(\{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}\}) = 3/4$ である。

# 離散型確率変数の例 (可算無限の場合)

Example

- $\Omega = \mathbb{N} = \{1, 2, \dots\}$.
- $P(\{n\}) = 1/2^n \quad \forall n \in \Omega$.

確率変数を $X(n) = n$ と定義すると $\Pr(X = n) = P(\{n\}) = 1/2^n$.

Example

- $\Omega = \{\mathrm{T}, \mathrm{HT}, \mathrm{HHT}, \mathrm{HHHT}, \mathrm{HHHHT}, \dots\}$.
- $P(\{H^n T\}) = p^n(1 - p) \quad \forall n \in \Omega$.

確率変数を $X(H^n T) = n$ と定義すると
$\Pr(X = n) = P(\{H^n T\}) = p^n(1 - p)$.

# 連続型確率変数の例

Example

- $\Omega = [0, 1)$.
- $P([a, b]) = b - a$ から定まる.

$X(\omega) = \omega$ とおくと、$\Pr(X \le a) = P(\{\omega \in \Omega \mid \omega \le a\}) = a$ である。

# 連続関数との合成

確率変数 $X\colon \Omega \to \mathbb{R}$ と $\mathbb{R}$ 上の連続関数 $f\colon \mathbb{R} \to \mathbb{R}$ について合成 $f \circ X\colon \Omega \to \mathbb{R}$ も確率変数となる。

このとき $\Pr(f \circ X \in O)$ と書く代わりに $\Pr(f(X) \in O)$ と書く。

例えば $\Pr(2X \geq 1.7)$, $\Pr(X + 1 \geq 3)$, $\Pr(X^2 < 2)$ のように書く。

また確率変数 $X\colon \Omega \to \mathbb{R}$ と $Y\colon \Omega \to \mathbb{R}$ の和 $X + Y$ もまた確率変数である。

# 二項分布

- $\Omega = \{\mathrm{H}, \mathrm{T}\}^n$.
- $\mathcal{F} = 2^\Omega$.
- $P(\{\omega\}) = 1/2^n \quad \forall \omega \in \Omega$.

確率変数 $X\colon \Omega \to \mathbb{R}$ を $X(\omega) =$「$\omega$ に含まれる H の数」と定義する。このとき、

$$\Pr(X = k) = \binom{n}{k} \frac{1}{2^n}.$$

ここで $P$ の定義だけ変更して、$X(\omega) = k$ となる $\omega$ について $P(\{\omega\}) = p^k (1-p)^{n-k}$ とすると、

$$\Pr(X = k) = \sum_{\omega \in X^{-1}(k)} P(\{\omega\}) = \binom{n}{k} p^k (1-p)^{n-k}$$

となる。

# 離散確率分布の例

- 二項分布: $n$ 回独立にコインを投げて $k$ 回表が出る確率
  $\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$.

- 幾何分布: 独立にコインを投げて $k$ 回目に初めて表が出る確率
  $\Pr(X = k) = (1-p)^{k-1} p$.

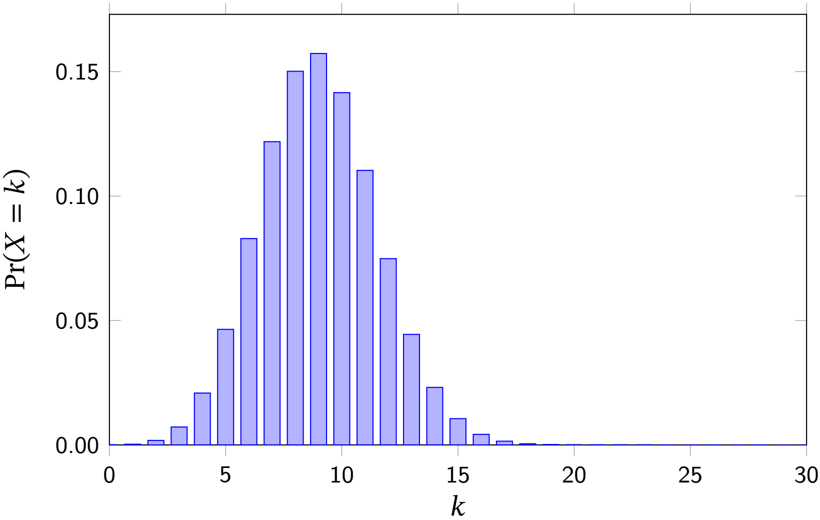
- 負の二項分布: 独立にコインを投げて $r$ 回表が出る前にちょうど $k$ 回裏が出る確率
  $\Pr(X = k) = \binom{r+k-1}{r-1} p^r (1-p)^k$.

- 超幾何分布: 袋の中に $N$ 個のボールがあって、そのうち $K$ 個が当たりのとき、$n$ 個引いて $k$ 個当たりを引く確率
  $\Pr(X = k) = \dfrac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$.

- ポアソン分布: $\Pr(X = k) = \dfrac{\lambda^k}{k!} \mathrm{e}^{-\lambda}$

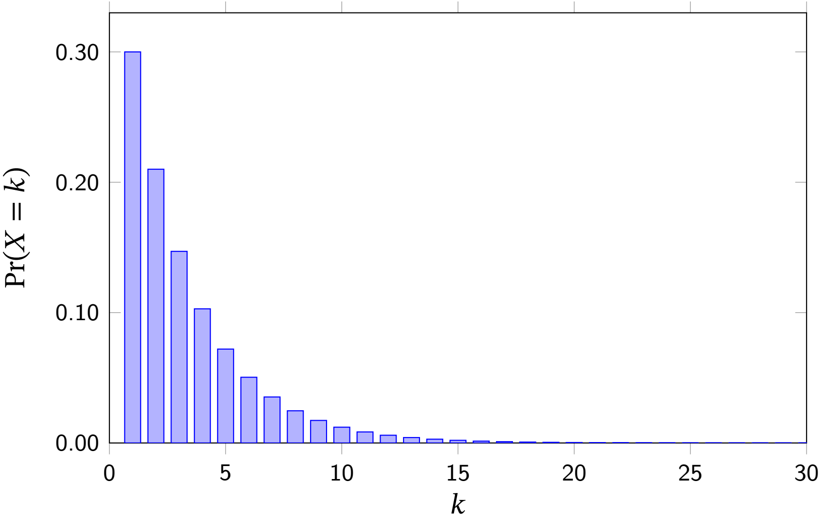
# 二項分布

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad n = 30,\ p = 0.3.$$



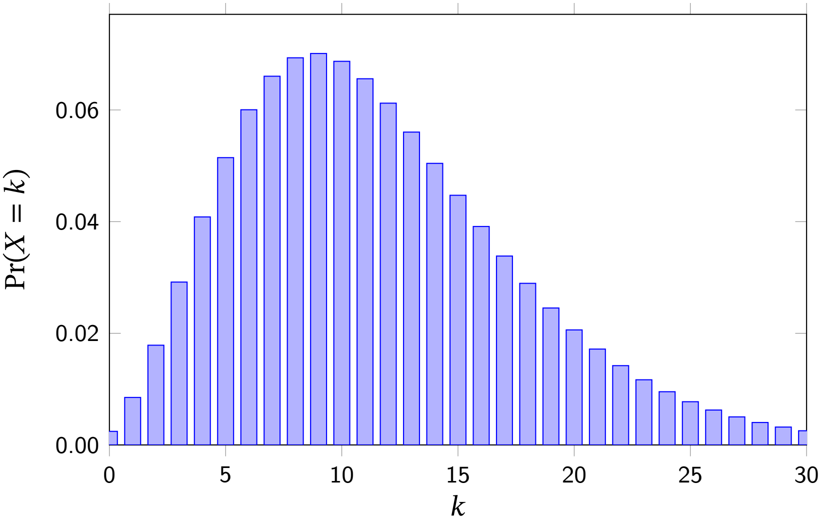
# 幾何分布

$$\Pr(X = k) = (1 - p)^{k-1} p, \quad p = 0.3.$$



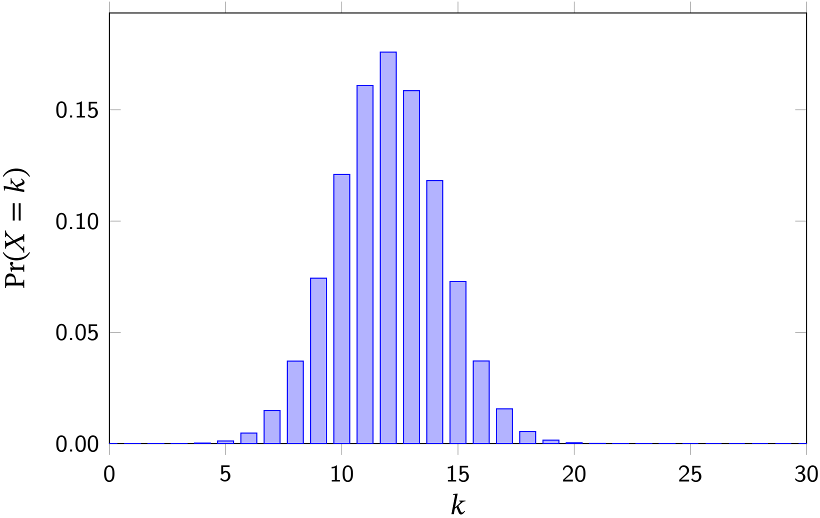
# 負の二項分布

$$\Pr(X = k) = \binom{r+k-1}{r-1} p^r (1-p)^k, \quad r = 5,\ p = 0.3.$$



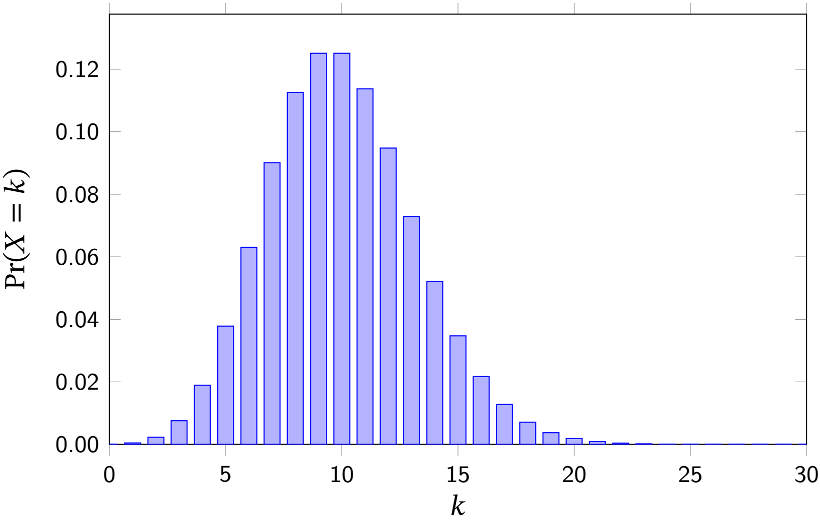
# 超幾何分布

$$\Pr(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, \quad N = 100,\ K = 30,\ n = 40.$$



# ポアソン分布

$$\Pr(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}, \quad \lambda = 10$$



# 同時確率と条件付き確率

離散確率変数 $X$ と $Y$ の同時確率

$$\Pr(X = x, Y = y)$$

周辺化

$$\Pr(X = x) = \sum_y \Pr(X = x, Y = y)$$
$$\Pr(Y = y) = \sum_x \Pr(X = x, Y = y)$$

条件付確率

$$\Pr(X = x \mid Y = y) = \frac{\Pr(X = x, Y = y)}{\Pr(Y = y)}$$

$Y = y$ のときの $X = x$ の確率 ($\Pr(Y = y) = 0$ のときは定義されない)

# 独立確率変数

離散確率変数 $X$ と $Y$ が独立

$$\overset{\text{def}}{\iff} \Pr(X = x, Y = y) = \Pr(X = x)\Pr(Y = y).$$

離散確率変数 $X$ と $Y$ が独立のとき

$$\begin{aligned}
\Pr(X \in A, Y \in B) &= \sum_{x \in A, y \in B} \Pr(X = x, Y = y) \\
&= \sum_{x \in A, y \in B} \Pr(X = x)\Pr(Y = y) \\
&= \Pr(X \in A)\Pr(Y \in B)
\end{aligned}$$

任意の連続関数 $f$ と $g$ について、
確率変数 $X$ と $Y$ が独立 $\implies$ $f(X)$ と $g(Y)$ は独立。

$$\begin{aligned}
\Pr(f(X) = x, g(Y) = y) &= \Pr(X \in f^{-1}(x), Y \in g^{-1}(y)) \\
&= \Pr(X \in f^{-1}(x))\Pr(Y \in g^{-1}(y)) \\
&= \Pr(f(X) = x)\Pr(g(Y) = y)
\end{aligned}$$

# 少数の法則

## Theorem (少数の法則)

非負実数 $\lambda \geq 0$ とする。非負整数 $n$ について $p_n = \lambda/n$ とする。任意に固定した非負整数 $k$ について、

$$\lim_{n\to\infty}\binom{n}{k}p_n^k(1-p_n)^{n-k} = \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}.$$

## Proof.

$$\begin{aligned}
\lim_{n\to\infty}\binom{n}{k}p_n^k(1-p_n)^{n-k} &= \lim_{n\to\infty}\binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \lim_{n\to\infty}\binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{-k}\left(1-\frac{\lambda}{n}\right)^{\frac{n}{\lambda}\lambda} \\
&= \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}.
\end{aligned}$$

□